The Aggregation Prejudice: Knowledge Compilation from Diderot to Wikipedia

In 1745, a Parisian publisher named André Le Breton hired Denis Diderot to translate an English encyclopedia into French. The source was Ephraim Chambers' *Cyclopaedia*, a respectable two-volume reference work that had sold well in London. What Le Breton wanted was a commercial product: a French edition for a French market. What Diderot produced, over the next twenty-five years, was something that neither Le Breton nor the French state had anticipated: a weapon.

The *Encyclopédie, ou dictionnaire raisonné des sciences, des arts et des métiers* appeared in its first volume in 1751. By its completion in 1772, it had grown to twenty-eight volumes: seventeen of text, eleven of plates. It contained 71,818 articles and 2,885 illustrations, the work of over 150 contributors, among them Voltaire, Rousseau, and Montesquieu. Its most prolific author was the Chevalier de Jaucourt, who wrote 17,266 articles, roughly eight per day across six years, a feat of sustained gathering that remains difficult to comprehend. Diderot himself wrote approximately seven thousand entries, spending days in workshops learning trades and nights recording what he had learned. He damaged his eyesight correcting proofs. He was imprisoned. His co-editor, Jean le Rond d'Alembert, resigned under pressure in 1758. The project consumed the central decades of Diderot's intellectual life.

The reception was immediate and violent. The first two volumes drew condemnation from the Archbishop of Paris, the Jesuits, and the Parlement, which suspended publication in 1752. In 1759, the French government formally suppressed the work; Pope Clement XIII condemned it and instructed Catholics to bring their copies to a priest for burning. Yet the *Encyclopédie* survived through unlikely allies. Chrétien-Guillaume de Lamoignon de Malesherbes, the royal censor (the man whose official duty was to suppress dangerous publications) warned Diderot of an impending police raid and hid the manuscripts in his own home. Madame de Pompadour, the king's mistress, intervened on the project's behalf. The later volumes were printed in secret, their title pages falsely claiming publication in Neuchâtel rather than Paris.

The *Encyclopédie*'s stated aim, as Diderot wrote in his own article on the word "Encyclopédie," was "to change the way people think": not by inventing new ideas, but by changing how existing ideas were arranged. His system of cross-references, the *renvois*, linked articles to one another, creating a web of knowledge that a reader could traverse in any direction. Some of these links were deliberately subversive: an orthodox article on a theological topic might refer the reader to a secular philosophical entry that quietly undermined the first. D'Alembert's "Preliminary Discourse," published with the first volume in 1751, mapped all human knowledge as a branching tree in which theology appeared as one branch among many, not the root. The synthesis Diderot achieved was not merely additive. The cross-reference system created relationships between fields that had not previously been placed in dialogue, and the classification itself constituted an argument about the structure of knowledge, even if no individual article said so.

What endangered Diderot was not originality in the conventional sense. Most articles in the *Encyclopédie* drew on existing knowledge: published scholarship, craft traditions, established science. The emergence of the work's true radicalism lay not in its content but in its architecture. Darnton (1979) has shown that the enterprise was, at its core, as much a commercial venture as an intellectual one. The danger lay in the gathering itself, in compilation as a political act. Scattered across specialist publications, guild workshops, and learned societies, this knowledge posed no threat. Bound together in one navigable structure, cross-referenced and indexed, accessible to anyone literate enough to read French, it became intolerable. The Church and the monarchy understood something that later intellectual history would largely forget: that the organisation of existing knowledge can be more dangerous than the production of new knowledge. Every channel of dissemination they controlled was bypassed by a single work of reference.

Even Diderot's own publisher turned against the project. In 1764, as the final volumes neared completion, Diderot discovered that Le Breton had been secretly censoring articles after Diderot had approved the proofs, altering roughly three hundred pages of material he considered too dangerous to print. Le Breton and his foreman then destroyed the original manuscripts to prevent the damage from being undone. In a letter of November 1764, Diderot accused Le Breton of having "massacred" the work. His confidant Friedrich Melchior Grimm described Le Breton as a "crazy murderer" who had left the *Encyclopédie* in "mutilated fragments." Diderot considered abandoning the project entirely. He did not. The final impression appeared in 1772.

The word "compiled" appears on the title page of the first edition of the *Encyclopaedia Britannica*. Published in Edinburgh between 1768 and 1771, the work announced itself as "A Dictionary of Arts and Sciences, Compiled upon a New Plan." It consisted of three volumes, roughly 2,500 pages, and 160 copperplate engravings. Its compiler, William Smellie, a twenty-eight-year-old printer, later joked that he had made it "with a pair of scissors," a self-deprecation that would follow the Britannica, and the encyclopedic tradition generally, for the next two and a half centuries.

The new plan was organisational: rather than listing terms alphabetically with brief definitions, as previous reference works had done, the Britannica gathered related topics into substantial treatises arranged within an alphabetical framework. The article on anatomy ran to 165 pages. The medium shaped the message: the longer treatise format allowed for sustained exposition that no mere dictionary entry could achieve. The emergence of this format marked a shift in ambition, from defining terms to teaching subjects.

Over succeeding editions, the Britannica evolved from compilation toward commissioning. The ninth edition (1875–1889), dubbed the "Scholar's Edition," recruited leading authorities in their fields. James Clerk Maxwell served as a special advisor on science; Thomas Huxley contributed on evolution. Roughly 1,100 contributors participated, a handful of them women. By the landmark eleventh edition of 1911, spanning twenty-nine volumes, over forty thousand articles, and more than 1,500 named contributors, the Britannica had become perhaps the most authoritative reference work in English. Shaw claimed to have read the entire ninth edition except the science articles. The eleventh is still cited in scholarly work more than a century after publication, its copyright having lapsed into the public domain. The Britannica had crossed from the substrate of one man's scissors to an institutional enterprise, its reception transformed from useful gathering to something

approaching secular canon.

Yet even at its most prestigious, the Britannica occupied an uneasy position in the hierarchy of intellectual labour. The scholars who wrote its articles received individual credit. The editors who selected, structured, and connected those articles, who decided what belonged in the canon and how knowledge should be traversed from entry to entry, did not. The work of synthesis was invisible by design: a well-functioning encyclopedia appears to present knowledge as it naturally is, not as someone has chosen to arrange it. The traversal it enables looks effortless precisely because the labour of construction has been erased.

The commercial model reinforced this ambiguity. Beginning in the early twentieth century, after American businessmen acquired the title, Britannica was sold door-to-door by a dedicated sales force that at its peak numbered over 2,300 representatives. Salesmen in living rooms pitched the encyclopedia as a pathway to education and social advancement, knowledge as domestic aspiration, bound in leather, delivered to the household on instalments. The channel through which knowledge reached families was personal solicitation, and the binding of the product (its physical weight, its shelf presence, its impression of permanence) was inseparable from its authority. Peak sales reached over 100,000 sets in 1990. By the time the company announced the discontinuation of its print edition in March 2012, after 244 years of continuous publication, print sales accounted for barely one percent of the business. The printed Britannica was not killed by a competitor. It was displaced by a different model of gathering operating on a different substrate, one that moved faster, cost nothing, and could be edited by anyone. The old channels had been routed around entirely.

Wikipedia launched on January 15, 2001, as a side project. Its predecessor, Nupedia, was a peer-reviewed online encyclopedia overseen by Larry Sanger and funded by Jimmy Wales through Bomis. Nupedia's seven-step review process, designed for professional quality, produced fewer than two dozen articles in its first year. Wikipedia was conceived as a way to break the bottleneck, a wiki-based platform where anyone could contribute drafts that might eventually feed into Nupedia's formal review process. Within days, the side project had more articles than the main one. Within months, Nupedia was an afterthought. It ceased operations in 2003, having completed only twenty-four articles through its full review process.

As of February 2026, the English Wikipedia contains over 7.1 million articles across 344 language editions. The site receives billions of page views per month and is edited, on average, nearly six times every second. The emergence of this corpus outpaced any individual's capacity to absorb it: reading every English article would take thirty-eight years.

The reception of Wikipedia by established intellectual institutions has followed a pattern that Diderot would have recognised. In 2005, the journal *Nature* published a study comparing the accuracy of forty-two science articles from Wikipedia and the *Encyclopaedia Britannica*. The reviewers found an average of 3.9 errors per Wikipedia article and 2.9 per Britannica article, a gap, certainly, but not the gulf that Wikipedia's critics had predicted. Britannica disputed the study's methodology; *Nature* stood by its findings. The episode was revealing less for its conclusions than for its framing: the question of whether Wikipedia was reliable enough had become, implicitly, a question about who had the authority to compile. The resistance was never primarily about

accuracy. It was about the channel through which knowledge was permitted to flow.

What distinguishes Wikipedia from its predecessors is not only its scale but the emergence of editorial labour into public view. The digital medium preserves what the print substrate erased: every article carries an edit history, a complete record of every addition, deletion, and revision. Talk pages document the debates behind editorial decisions. Deletion discussions, arbitration proceedings, and notability guidelines are all conducted in public. The work that was always present in compilation (the decisions about what to include, how to frame it, when to cut) is for the first time exposed. Wikipedia did not invent the labour of compilation. It made that labour legible. A reader's traversal of the encyclopedia could now accompany a traversal of the process that built it.

And still the charge persists. Academic style guides warn against citing Wikipedia. Instructors prohibit it as a source. The dismissal takes a familiar form: anyone can edit it; it is "merely" an aggregation of existing knowledge; it lacks the authority of expert judgment. These objections echo across centuries. They are the same objections that were raised against Diderot's *Encyclopédie* by the Jesuits, against the Britannica by academic specialists who considered it superficial, against every canon of compiled knowledge by those who believed that gathering was not the same as creating. Wales himself consistently described Wikipedia as an encyclopedia, not a social platform. The compilation model was explicit from the outset.

The inclusionist-deletionist debate, Wikipedia's ongoing internal argument about what belongs in the encyclopedia, is the compilation question stated in its purest form: who decides what to gather? For centuries, this decision was made in private: by Diderot and his contributors, by Britannica's editorial board, by the institutional gatekeepers of every reference work. On Wikipedia, the arguments happen on talk pages, in deletion discussions where editors marshal evidence for and against the notability of a given subject. For the first time in the history of knowledge gathering, this question is being negotiated in public, with every argument preserved. The emergence of this transparency has not, however, changed how the work is valued. The editors who maintain Wikipedia's millions of articles, correcting errors, mediating disputes, enforcing standards of evidence, remain largely anonymous, largely uncompensated, and largely uncredited. A small percentage of editors produce most content, with significant demographic skews in geography and gender. The infrastructure is vast, complex, and invisible to most of the people who use it daily.

The pattern is not subtle. A compiler gathers existing knowledge into a new structure. The emergence of that structure makes the knowledge accessible, navigable, useful, creates something that the scattered source material, on its own, could not provide. The compiler is then accused of "merely" gathering. The accusation comes most forcefully from those who produced the original knowledge and who see in the compiled work a threat to their authority, their gatekeeping function, their impression of how knowledge ought to move through the world.

Diderot's project took twenty-five years and was suppressed by the French crown, the Catholic Church, and the Parlement de Paris. Britannica took 244 years to build its authority and watched the collapse of its commercial model in barely a decade. Wikipedia assembled in years what earlier compilations had required centuries to achieve. The acceleration is itself part of the pattern: each new medium (print, institutional publishing, the open web) compresses the cycle while reproducing

its essential structure. The compiler gathers. The establishment resists. The emergence becomes indispensable. The compiler disappears into the work.

Consider what each compilation left behind. Diderot's cross-reference system, a web of knowledge that readers could traverse laterally, following connections the compiler had laid, anticipated the hyperlink by two centuries. Britannica's institutional model established the convention that an encyclopedia should commission rather than merely excerpt, binding expert authority to editorial structure in a form that shaped every reference work that followed. Wikipedia's edit histories and talk pages constitute the first comprehensive archive of how compiled knowledge is actually made, and yet this freely available archive remains largely unstudied outside a small community of information scientists. Reagle (2010) and Jemielniak (2014) have examined Wikipedia's collaborative culture, but the broader history of synthesis as intellectual practice has attracted remarkably little sustained scholarly attention.

What survives is never the compiler's name. Diderot is remembered as a philosopher and novelist; his editorial labour on the *Encyclopédie* is treated as biographical detail rather than his central intellectual achievement. The Britannica's editors are anonymous by convention; the institution is the author. Wikipedia's contributors are anonymous by design. In each case, the synthesis that the compiler performed, selecting, arranging, connecting, binding scattered knowledge into a coherent whole, vanishes into the product. The cross-references appear to have always existed. The articles appear to have written themselves. The navigable structure appears to be a property of knowledge rather than an artefact of someone's labour.

The compilation prejudice persists because it serves a function. If compilation is "merely" gathering, then the compiler deserves no credit, the knowledge appears to have organised itself, and the enormous labour of selection, synthesis, and structure can go uncompensated and unrecognised. The product remains. The process that created it (the decades in workshops, the editorial arguments, the midnight corrections of proof, the daily accumulation of small decisions about what belongs and what does not) is erased by its own success. The better the compilation, the more invisible the compiler. The work is always there. It is never seen.

Bibliography

Blom, Philipp. *Encyclopédie: The Triumph of Reason in an Unreasonable Age*. London: Fourth Estate, 2004.

Darnton, Robert. *The Business of Enlightenment: A Publishing History of the Encyclopédie, 1775–1800*. Cambridge, MA: Harvard University Press, 1979.

Giles, Jim. "Internet Encyclopaedias Go Head to Head." *Nature* 438 (2005): 900–901.

Jemielniak, Dariusz. *Common Knowledge? An Ethnography of Wikipedia*. Stanford: Stanford University Press, 2014.

Kafker, Frank A., ed. *Notable Encyclopedias of the Late Eighteenth Century: Eleven Successors of the Encyclopédie*. Oxford: Voltaire Foundation, 1994.

Kogan, Herman. *The Great EB: The Story of the Encyclopaedia Britannica*. Chicago: University of Chicago Press, 1958.

Reagle, Joseph. *Good Faith Collaboration: The Culture of Wikipedia*. Cambridge, MA: MIT Press, 2010.

Yeo, Richard. *Encyclopaedic Visions: Scientific Dictionaries and Enlightenment Culture*. Cambridge: Cambridge University Press, 2001.